PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-124097
(43)Date of publication of application : 06.05.1994

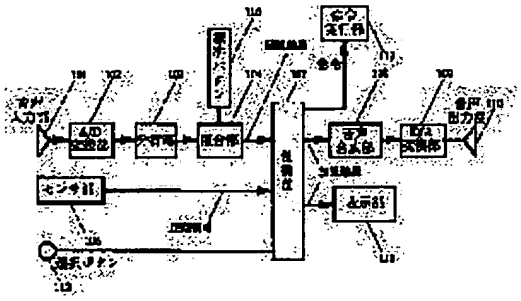
(51)Int.CI. G10L 3/00
G10L 3/00
G06F 3/16
G10L 3/02

(21)Application number : 04-274609
(22)Date of filing : 13.10.1992

(71)Applicant : HITACHI LTD
(72)Inventor : KOKUBO HIROAKI

(54) PORTABLE TERMINAL DEVICE

(57)Abstract:
PURPOSE: To enable a user to select proper usage according to the purpose and use environment by securing a sufficient S/N ratio even in a noisy atmosphere when operating the portable terminal device near the mouth and confirming a recognition result by voice without looking at a display, and confirming recognition candidates on the display when using the device at a distance from the mouth.
CONSTITUTION: This portable terminal device has a voice recognizing function and is equipped with a voice input part 101 which inputs a voice to be recognized, voice recognition part 103-105 which recognize the input voice obtained by the voice input part 101, a sensor part 106 which measures the distance between the input part 101 and user, a voice synthesis part 108 which synthesizes a voice basing upon at least the recognition results of the voice recognition parts 103-105, a voice output part 110 which reproduces the synthesized voice, a display part 111 which displays the recognition results, and a control part 107 which displays the recognition results through the voice output part 110 when the distance measured by the sensor part 106 is smaller than a previously defined threshold value or at a display part 111 when not.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

1. This document has been translated by computer.So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

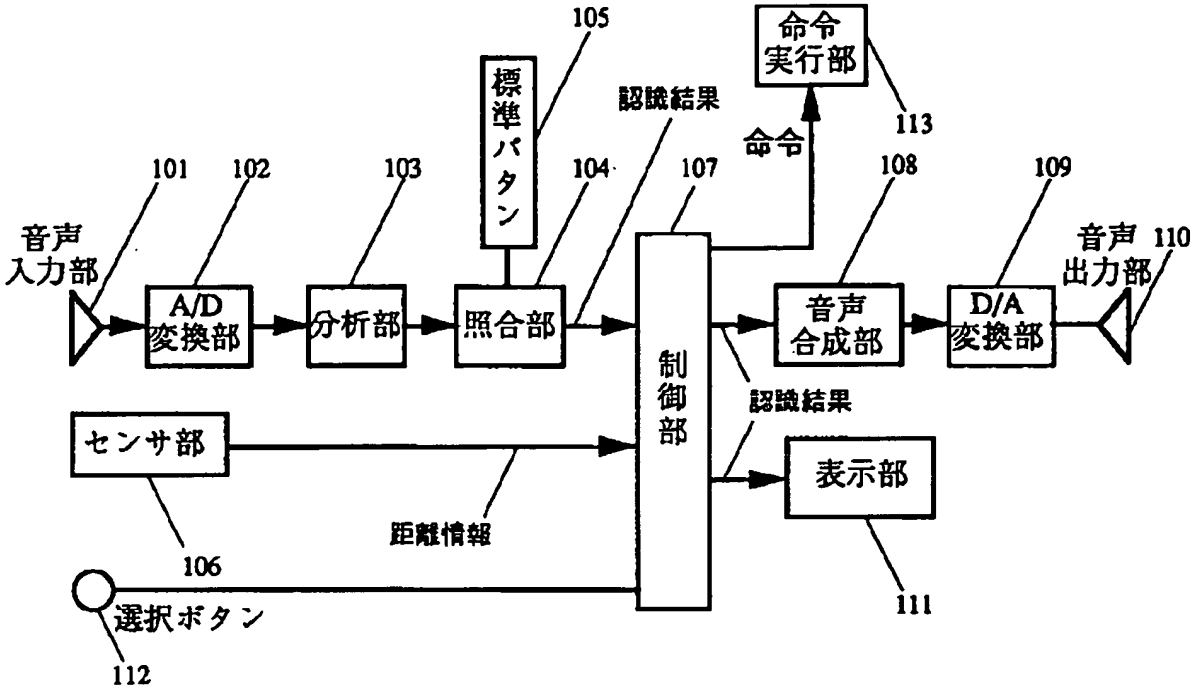
[Claim(s)]
[Claim 1] Personal digital assistant equipment equipped with the speech recognition function characterized by providing the following. The voice input section which inputs the voice used as the candidate for recognition. The speech recognition section which recognizes the input voice obtained from this voice input section. The sensor section which measures the distance of the aforementioned voice input section and a user. The control section show a recognition result from the aforementioned voice output section when the distance measured in the speech-synthesis section which compounds the voice based on the recognition result in the aforementioned speech-recognition section at least, the voice output section for reproducing the this compounded voice, the display for displaying the aforementioned recognition result, and the aforementioned sensor section is small than the threshold defined beforehand, and show a recognition result to the aforementioned display in the aforementioned distance's being more than the aforementioned threshold.
[Claim 2] The carried type terminal unit according to claim 1 characterized by preparing two or more aforementioned sensor sections in a different place.
[Claim 3] The carried type terminal unit according to claim 1 characterized by changing the gain of the aforementioned voice input section according to the distance measured in the aforementioned sensor section.
[Claim 4] The carried type terminal unit according to claim 1 characterized by changing the volume of the aforementioned voice output section according to the distance measured in the aforementioned sensor section.
[Claim 5] It is the carried type terminal unit according to claim 1 characterized by removing noise in the aforementioned noise processing section when the distance which prepared the noise processing section which removes noise by carrying out frequency analysis of the aforementioned input signal, and subtracting a presumed noise spectrum to this voice spectrum by which frequency analysis was carried out, and was measured in the aforementioned sensor section is larger than the threshold defined beforehand.
[Claim 6] The carried type terminal unit according to claim 5 characterized by adjusting the size of the normal mode rejection in the aforementioned noise processing section according to the size of the distance measured in the aforementioned sensor section.
[Claim 7] The carried type terminal unit according to claim 1 or 5 characterized by using the standard pattern created using the aforementioned noise superposition data when the distance which prepares at least one kind of standard pattern created in the aforementioned speech recognition section using the noise superposition data according to the noise environment assumed beforehand, and was measured in the aforementioned sensor section is larger than the threshold defined beforehand.
[Claim 8] The aforementioned recognition result is a carried type terminal unit according to claim 1 characterized by showing two or more recognition candidates when only the 1st place only of a candidate is shown when showing from the aforementioned voice output section, and the aforementioned display shows.
[Claim 9] It is the carried type terminal unit according to claim 1 characterized by the aforementioned instruction-execution section performing processing according to the voice instruction of the recognition candidate of the 1st place when there is no fixed time input after having the instruction-execution section which performs processing according to the voice instruction recognized in the aforementioned speech recognition section and showing a recognition candidate from the aforementioned voice output section.
[Claim 10] The carried type terminal unit according to claim 1 characterized by using the size of the input level of the voice inputted from the aforementioned voice input section instead of using the value measured by the aforementioned distance robot.
[Claim 11] The carried type terminal unit according to claim 1 characterized by having the switch which chooses whether a recognition result is outputted with voice, or it displays on a display instead of measuring distance by the aforementioned sensor.

[Translation done.]

DRAWINGS

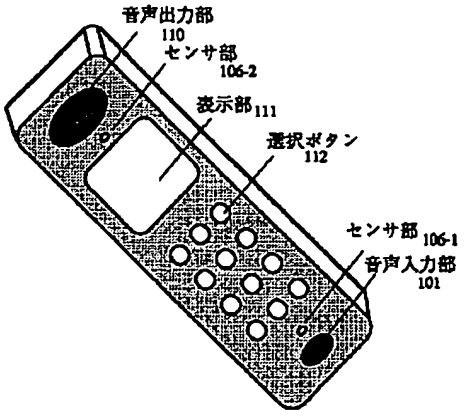
[Drawing 1]

システム構成図(図1)



[Drawing 2]

外観構成図(図2)



[Drawing 5]

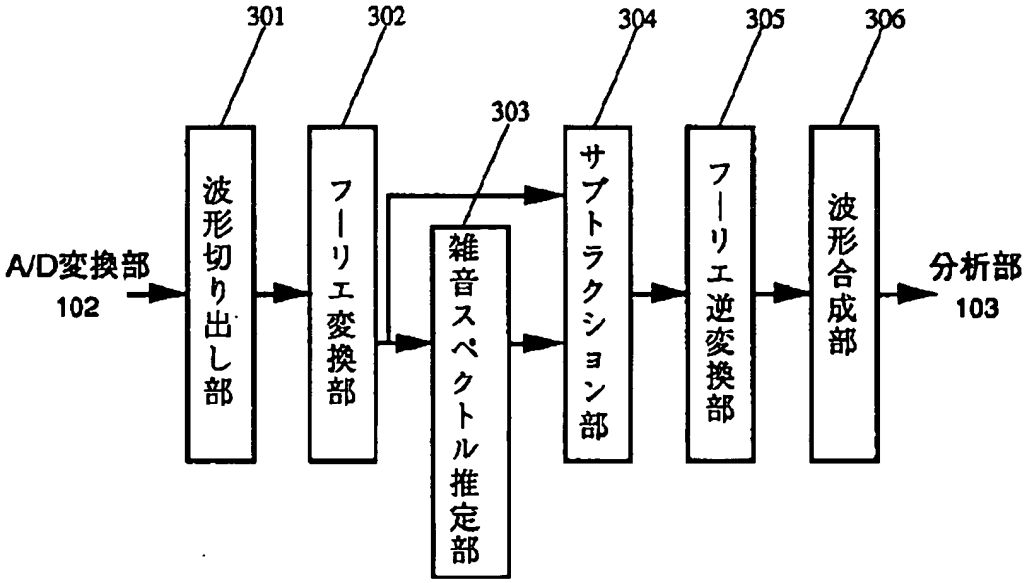
音声入力に対する出力例（図5）

音声入力「なかた」

	1001 認識結果	1002 音声出力例	1003 表示部出力例
ケース1	<リジェクト>	「もう一度お願いします。」	もう一度お願いします。
ケース2	1．なかた	「"なかた"さんに電話をします。」	"なかた"さんに電話します。 1 はい 0 いいえ
ケース3	1．なかた 2．なかた 3．たかだ	「"なかた"さんに電話をします。」	電話したい人を選んでください。 1 なかた 2 なかた 3 たかだ 0 取消

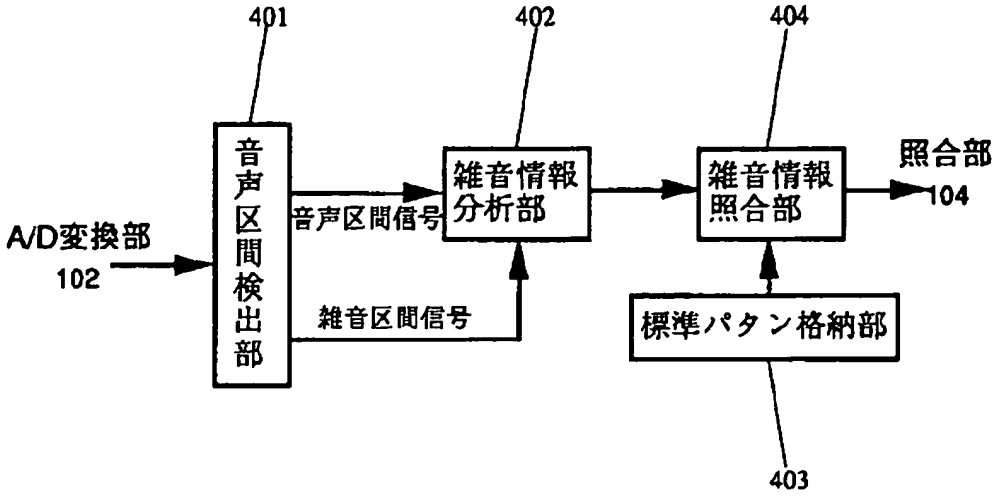
[Drawing 3]

雑音処理部 (図3)



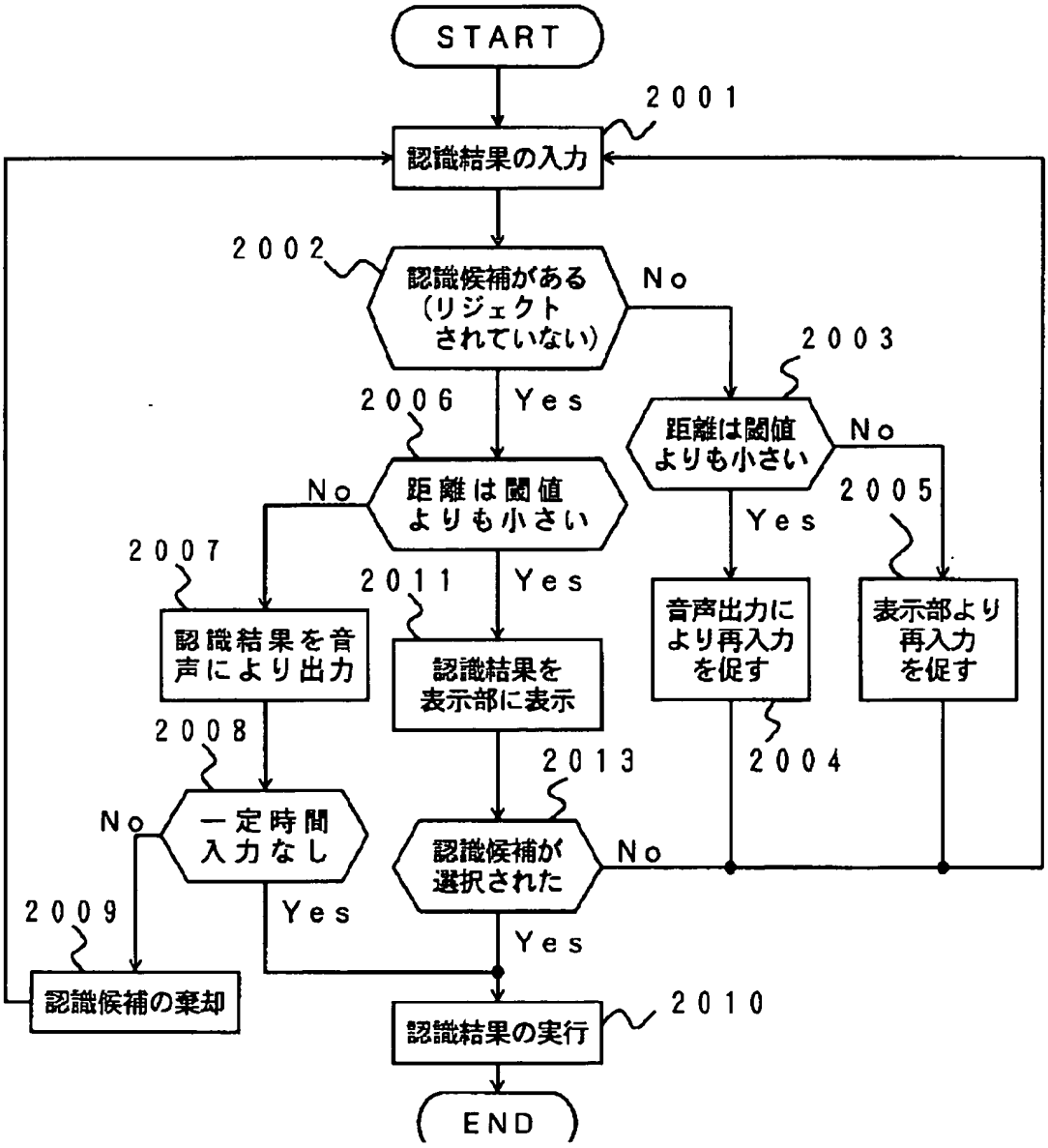
[Drawing 4]

標準パタン選択部 (図4)



[Drawing 6]

制御部の処理フロー（図6）



[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram showing the composition of one example of this invention.
[Drawing 2] It is the external view of the telephone in an example.
[Drawing 3] It is the block diagram showing the composition of the noise processing section which can be added to the composition of drawing 1 .
[Drawing 4] It is the block diagram showing the composition of the standard-pattern selection section which can be added to the composition of drawing 1 .
[Drawing 5] It is explanatory drawing of the example of an output to the voice input in an example.
[Drawing 6] It is the flow chart which shows the processing flow of a control section shown in drawing 1 .
[Description of Notations]
101 [-- The analyzor, 104 / -- The collating section, 105 / -- A standard pattern, 106 / -- The sensor section, 107 / -- A control section, 108 / -- The speech synthesis section, 109 / -- The D/A-conversion section, 110 / -- The voice output section, 111 / -- A display, 112 / -- Selection section.] -- The voice input section, 102 -- The A/D-conversion section, 103

[Translation done.]

* NOTICES *

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] this invention relates to the carried type terminal unit equipped with the speech recognition function, especially relates to usable carried type terminal unit equipment also under noise.
[0002]
[Description of the Prior Art] In the small carrying type terminal unit represented by the cellular phone, the number and size of small, therefore an operation button are limited. Therefore, operation using a handwriting character, voice, etc. is desired rather than it operates it with a button to such equipment.
[0003] The troubles of speech recognition are that it not necessarily always cannot necessarily recognize correctly because of the ambiguity which voice has, and a point that a recognition performance will deteriorate sharply for circumference noise if it is used under noise environment. Especially in recognition of a large vocabulary, or recognition with many similar words, the former becomes with a problem. In order to prevent the operation mistake by such recognition error, even when it recognizes accidentally, it is required to show the next candidate and to carry out error recovery well.
[0004] Moreover, in order to put the carried type terminal unit equipped with the speech recognition function in practical use, the technology of noise[-proof]-izing which can be correctly recognized also with the voice uttered under noise is also indispensable. The technique of removing noise from the voice which noise superimposed by pretreatment as a means to make the voice uttered under noise environment recognize correctly, or the recognition technique which can be correctly recognized also with the voice superimposed on noise is required. The former presumes what removes noise using an adaptation filter, and the noise spectrum mixed in voice, and has the spectrum subtraction technique deducted from an input spectrum. There are technique using the parameter and interval scale which cannot be easily influenced of noise, the noise superimposing method which superimposes noise on the standard pattern beforehand in the latter. However, although much noise processing technique is proposed, as compared with the recognition performance under still quiet environment, it cannot say that it is enough, but the best cure against noise is raising S/N of input voice if possible using a close-talking microphone etc.
[0005]
[Problem(s) to be Solved by the Invention] In order for the technique of checking the recognition result displayed on the display in order to mistake and to check a recognition result that there is nothing, and the technique which a user is made to choose from two or more recognition candidates to be able to recognize correctly also under noise environment desirably, it is required to make distance of a mouth and a microphone small if possible, and to raise S/N like a close-talking microphone. it is alike, and an appropriate thing [making a microphone and a display separate without spoiling portability and user-friendliness in a small carrying type terminal unit] is next to impossible, and it is difficult to check a display by the eye in the case where it is used bringing a microphone close to the month When operating it in the position which is distant from the month on the other hand in order to check a recognition result by the display, it is difficult to be unable to secure sufficient S/N but to secure a satisfying recognition performance in a place with much noise.
[0006] It is in the ability of a user to be made for proper use to do, corresponding to the purpose or an operating environment so that operation may be possible, being able to bring close to the month when the improvement in a recognition rate gives priority also to the easy halfbeak of check operations [place] in the carrying type terminal unit whose purpose of this invention was equipped with the speech recognition function, being able to operate it, separating from the month in the place where it is comparatively few, and checking a recognition result by the display
[0007]
[Means for Solving the Problem] In order to attain the above-mentioned purpose, the carried type terminal unit by this invention The voice input section which is personal digital assistant equipment equipped with the speech recognition function, and inputs the voice used as the candidate for recognition, The speech recognition section which recognizes the input voice obtained from this voice input section, and the sensor section which measures the distance of the aforementioned voice input section and a user, The speech synthesis section which compounds the voice based on the recognition result in the aforementioned speech recognition section at least, The voice output section for reproducing the compounded this voice, and the display for displaying the aforementioned recognition result, It is characterized by having the control section which presents a recognition result from the aforementioned voice output section when the distance measured in the aforementioned sensor section is smaller than the threshold defined beforehand, and shows the aforementioned display a recognition result when the aforementioned distance is more than the aforementioned threshold.

[0008]
[Function] Although many deformation can be considered to this invention, the operation is explained about a typical means in it.
[0009] It sets beforehand, and in being smaller than a **** threshold, from the voice output section, the distance measured in the sensor section presents a recognition result, and, in more than a threshold, presents a recognition result conversely at a display. When operating it by the month, even if it can secure S/N sufficient also in the bottom of noise and does not see a display (display) by this, a recognition result can be checked with voice. Moreover, since the check of the recognition candidate by the display is possible when using it, separating from the month, even when there is a similar recognition candidate, a user can choose. Therefore, by using properly according to the purpose or an operating environment, in a place with much noise, S/N sufficient by bringing close to the month and operating it is secured, a high recognition performance is obtained, and comparatively, a user can separate from the month and can check two or more recognition candidates by the display in the few place of noise. For this reason, recognition of a large vocabulary and the recognition in which many similar words are included are also easy error recovery. For this reason, operation can be carried out, without hardly producing the stress by the recognition error.
[0010]
[Example] Hereafter, a drawing explains the example of this invention in detail.
[0011] It is the block diagram showing the composition of one example of the carried type terminal unit which equipped drawing 1 with the speech recognition function by this invention. this example explains taking the case of a cellular phone, as the appearance is shown in drawing 2 . However, this invention can be applied also like small personal digital assistants, such as an electronic notebook and remote control equipment, in addition to a cellular phone.
[0012] drawing 1 and drawing 2 -- setting -- 101 -- the voice input section and 102 -- the A/D-conversion section and 103 -- the analyzor and 104 -- the collating section and 105 -- a standard pattern and 106 -- the sensor section and 107 -- for the D/A-conversion section and 110, as for a display and 112, the voice output section and 111 are [a control section and 108 / the speech synthesis section and 109 / a selection button and 113] the instruction-execution sections The voice input section 101 is a portion which inputs voice, such as a voice command. It is quantized by the A/D-conversion section 102 and the sound signal inputted from the voice input section 101 goes into the analyzor 103. In the analyzor 103, the feature vector of the voice used for the judgment of recognition using well-known tools of analysis, such as LPC analysis, is extracted. About the audio feature extraction method, it is detailed to Furui "digital speech processing" Tokai University Press etc. To the feature vector extracted by the analyzor 103, in the collating section 104, it asks for a recognition candidate by calculating the degree of similar with the standard pattern 105 which consists of the feature vector of the vocabulary for recognition, and he is outputted to a control section 107, using the recognition candidate of a high order as a recognition result. The collating section 104 outputs a rejection signal as a recognition result instead of a recognition candidate, when constant value with the 1st place of a recognition candidate's degree of similar is not satisfied.
[0013] The sensor section 106 is for measuring the distance of a user and equipment, and can be easily realized by using distance robots, such as an infrared sensor and an ultrasonic sensor. Although it is sufficient for the sensor section 106 if the one function top exists, it is desirable to install in at least two positions mutually left as shown in drawing 2 in consideration of possibility of covering when had in a hand.
[0014] A control section 107 controls the method of showing a recognition result etc. using the distance information searched for in the sensor section 106. That is, when smaller than a threshold with the distance value inputted into the control section 107, a recognition result is passed to the speech synthesis section 108, and a display 111 is passed when distance is conversely larger than a threshold.
[0015] First, the case where the inputted distance is small is explained previously. A control section 107 passes a recognition result to the speech synthesis section 108, and compounds the voice of the recognition candidate of the 1st place. Moreover, when a rejection signal is inputted into the speech synthesis section 108, the voice of the guidance to which reinput is urged is compounded. With guidance voice, it says, "I need your help once again." If the voice data for reproduction is stored of course beforehand, the speech synthesis section is unnecessary. After the voice compounded in the speech synthesis section 108 is changed into an analog signal by the D/A-conversion section 109, it is reproduced from the voice output section 110. When the recognition result newer than the collating section 104 has inputted after a control section 107 passes a recognition result to the speech synthesis section 108, it rejects the past recognition result and updates a recognition result. Moreover, if there is no input of fixed time voice after outputting a recognition result, a control section 107 will output a recognition result to the instruction-execution section 113. Thereby, a user becomes possible [reinputting a voice command] from the voice input section 101 again, when the recognition candidate who outputted from the voice output section 110 is mistaken, or when it is the guidance voice which gives reinput. The instruction-execution section 113 is a portion which executes a voice instruction, for example, the dial section of voice dialing is equivalent to this.
[0016] Next, the case of a value with a larger distance inputted into the control section 107 than a threshold is explained. A display 111 is for displaying the recognition result inputted from the control section 107 as alphabetic information. The recognition result shown by the display 111 can also show simultaneously two or more recognition candidates similar also besides showing only the 1st place only of a candidate. For example, even the 3rd place even of a high order with the large degree of similar can be shown, or the application which is beyond a value with the degree of similar for which it asked in the collating section 104 of carrying out thing presentation can be considered. The check to the recognition result shown by the display 111 is performed by the selection button 112. For example, only in the case of a candidate, by pushing a confirmation

button or a cancellation button, the 1st place of the shown recognition result is by choosing the button matched with each candidate, when two or more candidates are shown. In addition, what does not necessarily need to be a button if the function is equipped with the same function, for example, attached the touch sensor to the display 111 is sufficient as the selection button 112. Moreover, a user can also input a voice command again from voice input 101 instead of choosing a button. A control section 107 outputs a recognition result to the instruction-execution section 113, when the right recognition candidate decides with the selection button 112.

[0017] Thus, even if according to this example it shows a recognition result from the voice output section when equipment is brought close to a face, it can secure S/N sufficient also in the bottom of noise when operating it by the month, in order to show a display a recognition result, when it separates from a face, and it does not see a display, a recognition result can be checked with voice. Moreover, even when there is a recognition candidate who was similar on a display since the check of a recognition candidate was possible when using it, having separated from the month, a user can choose the right candidate. Therefore, since S/N sufficient by bringing close to the month in a place with much noise, and operating it is secured, a high recognition performance can be obtained, and a user can be comparatively separated from the month in the few place of noise and can check two or more recognition candidates by the display by using properly according to the purpose or an operating environment, recognition of a large vocabulary and the recognition in which many similar words are included are also easy for error recovery. For this reason, operation can be carried out, without hardly producing the stress by the recognition error.

[0018] Although S/N is improved by bringing the voice input section 101 close to the month and improvement in the recognition rate under noise environment is aimed at in the above-mentioned example, it cannot be overemphasized that improvement in the recognition rate of the grade which is also the case where separate from the month and it is operated by preparing the noise processing section before the analyzor 103 can be aimed at. An example of the composition of the noise processing section is shown in drawing 3 . In this noise processing section, noise is removed using the technique called spectrum subtraction technique. About a spectrum subtraction, it is detailed to S.F.Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trsns.onAcoustics, Speech, and Signal processing, Vol.ASSP-27, No.2, April 1979, and pp.113-120. For a Fourier transducer and 303, as for the subtraction section and 305, in drawing 3 , the noise spectral estimation section and 304 are [301 / the wave logging section and 302 / a Fourier reverse transducer and 306] the wave composition sections. The digital signal outputted from the A/D-conversion section 102 is inputted into the wave logging section 301. The wave logging section 301 starts the wave section for analyzing spectrum information from an input signal, and starts the section for about dozens of ms at a fixed interval. The cut-down section signal wave type is changed into spectrum data in the Fourier transducer 302. Here, after hanging the windowing function used for the started wave conventionally [, such as a humming aperture,], a fast Fourier transform can be performed by embedding zero data in order and considering as the data of the factorial mark of 2, and high-speed data processing is realized. The spectrum signal by which the Fourier transform was carried out is inputted into the noise spectral estimation section 303. The noise spectral estimation section 303 calculates PAWA of a section spectrum signal, considers that the section where the value of the PAWA is less than a threshold more than fixed time is the non-voice section, and presumes a noise spectrum using the spectrum signal of the section. In addition to this, a lot of technique is proposed by the method of detecting the non-voice section (voice section), and it is also possible to detect the non-voice section using those technique. Although some are considered also about a noise spectrum estimation method using the signal of the non-voice section, the average spectrum of the spectrum for several frames is calculated and presumed, for example. In the subtraction section 304, a spectrum is subtracted using the noise spectrum presumed in the noise spectral estimation section 303 to the input spectrum signal. A subtraction is expressed with the following formula when X (f) and a presumed noise spectrum are now set to N (f) for the spectrum of input voice.

[0019]

[Equation 1]

$$\hat{S}(f) = \begin{cases} \dfrac{|X(f)| - \alpha \cdot |N(f)|}{|X(f)|} & \text{if } |X(f)| \geq \alpha |N(f)| \\ 0 & \text{if } |X(f)| < \alpha |N(f)| \end{cases} \quad \cdots\cdots\cdots \text{(数1)}$$

[0020] alpha in several 1 is called subtraction coefficient, and the effect of a normal mode rejection becomes large, so that this value is enlarged. However, since it will be removed to a voice component if alpha is enlarged too much, cautions are required for selection of a value. In this example, it is possible to judge that S/N is getting worse, so that the distance measured in the sensor section 106 is large, and to enlarge the value of alpha. Moreover, although it is subtracting to the amplitude of a spectrum in several 1, it is also possible to subtract by subtracting or putting in a phase component using a power spectrum.

[0021] The spectrum which removed the noise component in the subtraction section 304 is again changed into the signal of a time domain by the Fourier reverse transducer 305, and the frame data point currently started per frame is again compounded as a voice wave of a basis in the wave composition section 306, and is outputted to the analyzor 103. If the frame period of the noise processing section and the frame period of the analyzor 103 are made in agreement, of course, it is possible to output frame data to the analyzor 103 as it is, without using the wave composition section 306.

[0022] Moreover, two or more kinds of standard patterns doubled with noise environment are prepared, and choosing according to an operating environment can also improve the recognition rate under noise environment. An example of the

composition of the standard-pattern selection section is shown in drawing 4 . For 401, as for the noise information-analysis section and 403, in drawing 4 , a voice section detecting element and 402 are [the standard-pattern storing section and 404] the noise information collating sections. The voice data used for standard-pattern creation uses what superimposed noise on voice according to the operating environment of recognition equipment. The normal-mode-rejection signal outputted from the noise processing section 106 is divided into a voice section signal and a noise section signal in the voice section detecting element 401. In the noise information-analysis section 402, the inputted noise component is analyzed and the analysis parameter is outputted. The standard-pattern storing section 403 has stored the characteristic quantity of the noise component which superimposed the noise from which the kind differed on some kinds of standard patterns created from the voice data superimposed, respectively, and the voice data used for creation. The characteristic quantity of a noise component uses the same tools of analysis as what was used in the noise information-analysis section 402. The noise information collating section 404 collates the characteristic quantity of the noise component of a noise processing signal, and the characteristic quantity of the superposition noise stored in the standard-pattern storing section 403, and chooses and outputs the standard pattern created using the voice which superimposed the nearest noise on the noise component of a noise processing signal out of the standard pattern which is in the standard-pattern storing section 403 from the collating result. Moreover, the noise information-analysis section 402 can also calculate S/N of input voice instead of asking for the feature pattern of noise. Moreover, as Japanese Patent Application No. No. 329063 [three to] which these people proposed previously described, it is also possible to prepare both the noise processing section and the standard-pattern selection section.

[0023] Thus, even when detaching and using it from a face for the bottom of noise environment by preparing the noise processing section and two or more kinds of standard patterns, the recognition error by noise can be decreased. High recognition of noise-proof [nearby] nature is attained from what used each independently by these.

[0024] As mentioned above, in the collating section 140, the degree of similar makes the thing more than a threshold a recognition candidate, it inputs into a control section 107, and a control section 107 changes an output with voice, and the output by the display at any time based on the distance information acquired by the sensor section 106. Here, each example of an output of voice and a display is explained taking the case of voice dialing in telephone.

[0025] The example of an output of the voice output to voice input and a display output is shown in drawing 5 . In this drawing, the recognition result 1001 is as a result of [which is inputted into a control section 107 from the collating section 104 to voice input] recognition. The example 1002 of a voice output shows the example of an output in the case of outputting as voice, and the example 1003 of a display output shows the example of an output in the case of outputting as a display.

[0026] It divides into three cases and these examples of an output are explained. A case 1 is the case where a recognition candidate was not found but a rejection signal occurs. In this case, by the voice output, it outputs saying, "I need your help once again", and reinput is demanded from a user by displaying, "I need your help once again" with a display output. A case 2 shows the case where the number of recognition candidates is one ("backlash"). in this case -- when outputting saying, ""Mr. backlash" is telephoned" when outputting with voice, and outputting to a display, while displaying, ""Mr. backlash" is telephoned" -- alternative "-- 1 is and displays " and no ["no / 0 /"] A case 3 shows the case ("it is in **") where there are two or more recognition candidates. ["backlash", "inside **", and] In this case, when outputting with voice, only the 1st place only of a recognition candidate is employed, and it outputs, saying, ""Mr. backlash" is telephoned." the time of on the other hand displaying on a display -- "please choose those who want to telephone" -- displaying -- two or more selection candidate "1 -- backlash", "** in 2 ", "it being in 3 **", and "0 cancellation" are displayed

[0027] In addition, the example of an output explained above is a thing for explanation only, and this invention is not limited to this. Moreover, although explained taking the case of voice dialing, it is applicable also to uses other than voice dialing.

[0028] Next, the processing flow of a control section 107 is shown in drawing 6 . A control section 107 demands reinput from a user, when it judges whether the inputted recognition result is a recognition candidate or it is a rejection signal (2002) and a recognition candidate does not exist, if a recognition result is inputted from the collating section 104 (2001). As compared with the threshold which was able to define this beforehand, it judges whether reinput is guided with voice, or it carries out by display using the distance information specified in the sensor section 106 (2003). When distance is smaller than a threshold, a rejection signal is sent to the speech synthesis section 108, and reinput is urged by the voice output (2004). When distance is more than a threshold, a rejection signal is sent to a display 111 and reinput is urged from a display (2005).

[0029] On the other hand, when a recognition candidate exists, similarly, distance is compared with a threshold and a recognition candidate's presentation method is changed.

[0030] First, the case where a recognition result is outputted with voice is considered. After outputting a recognition result with voice (2007), a control section 107 waits to input a recognition result again from the collating section 104 fixed time (2008). When a recognition result is inputted again, while judging that the recognition candidate who showed was mistaken and rejecting a recognition candidate (2009), it performs from processing 2002 to a new recognition recognition result. Moreover, when there is no fixed time input, it judges that the recognition candidate who showed is right, and a recognition result is performed (2010).

[0031] Next, the case where a recognition candidate is outputted to a display 111 is considered. In this case, a recognition candidate is displayed on a display 111 with the guidance to which selection is urged (2011). If a recognition candidate is chosen with the selection button 112 (2013), the selected recognition result will be performed (2010). Moreover, when a recognition result is not chosen by cancellation button selection etc., it returns to processing 2001 again and waits (2013) and to input a recognition result again.

[0032] As mentioned above, although the example changed completely has explained the presentation method of a

recognition result bordering on the threshold of distance with a face in this example, when distance with a face is near the threshold, it is also possible to show both simultaneously. Moreover, it is also possible to make adjustable the gain of the voice input section 101 and the output level of the voice output section 110 according to distance with a face. Moreover, it is also possible as deformation of this example to use the size of the voice inputted into the voice input section 101 instead of measuring distance in the sensor section 106. That is, since the audio input level is in inverse proportion to the square of the distance of a mouth and the voice input section 101, the same effect is expectable if a threshold can be set up well. Moreover, it is also possible to use the switch with which a user can choose similarly whether a recognition result is outputted with voice or it displays on a display. Since the above deformation does not need the sensor section 106, its restrictions on a design decrease and it also has the advantage that a manufacturing cost can be made cheap.

[0033]

[Effect of the Invention] When [which was described above] operating it by the month, even if it can secure S/N sufficient also in the bottom of noise and does not see a display like according to this invention, a recognition result can be checked with voice. Moreover, even when there is a recognition candidate who was similar on a display since the check of a recognition candidate was possible when using it, having separated from the month, a user can choose the right candidate. Therefore, since S/N sufficient by bringing close to the month in a place with much noise, and operating it is secured, a high recognition performance can be obtained, and a user can be comparatively separated from the month in the few place of noise and can check two or more recognition candidates by the display by using properly according to the purpose or an operating environment, recognition of a large vocabulary and the recognition in which many similar words are included are also easy for error recovery. For this reason, operation can be carried out, without hardly producing the stress by the recognition error.

[Translation done.]